ABSTRACT

A technique that uses a weighted divide and conquer approach for clustering a set S of n data points to find k final centers. The technique comprises 1) partitioning the set S into P disjoint pieces $S_1,...,S_P$; 2) for each piece S_i, determining a set D_i of k intermediate centers; 3) assigning each data point in each piece S_i to the nearest one of the k intermediate centers; 4) weighting each of the k intermediate centers in each set D_i by the number of points in the corresponding piece S_i assigned to that center; and 5) clustering the weighted intermediate centers together to find said k final centers, the clustering performed using a specific error metric and a clustering method A.